CHAPTER

1

VECTORS AND MATRICES

Linear algebra provides a beautiful example of the interplay between two branches of mathematics, geometry and algebra. Moreover, it provides the foundations for all of our upcoming work with calculus, which is based on the idea of approximating the general function locally by a linear one. In this chapter, we introduce the basic language of vectors, linear functions, and matrices. We emphasize throughout the symbiotic relation between geometric and algebraic calculations and interpretations. This is true also of the last section, where we discuss the determinant in two and three dimensions and define the cross product.

1 VECTORS IN $\mathbb{R}^n$

A point in $\mathbb{R}^n$ is an ordered n-tuple of real numbers, written $(x_1, \dots, x_n)$. To it we may associate the vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, which we visualize *geometrically* as the arrow pointing from the origin to the point. We shall (purposely) use the boldface letter $\mathbf{x}$ to denote both the point and the corresponding vector, as illustrated in Figure 1.1. We denote by $\mathbf{0}$ the vector all of whose coordinates are 0, called the *zero vector*.

More generally, any two points A and B in space determine the arrow pointing from A to B, as shown in Figure 1.2, again specifying a vector that we denote $\overrightarrow{AB}$. We often refer

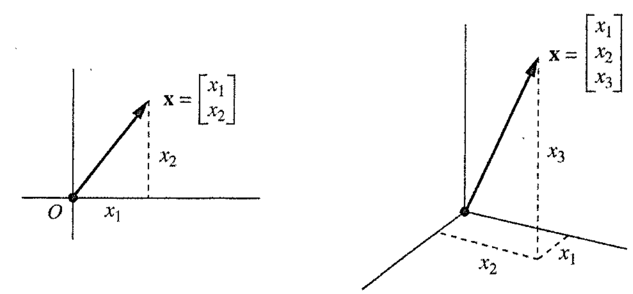


Figure 1.1

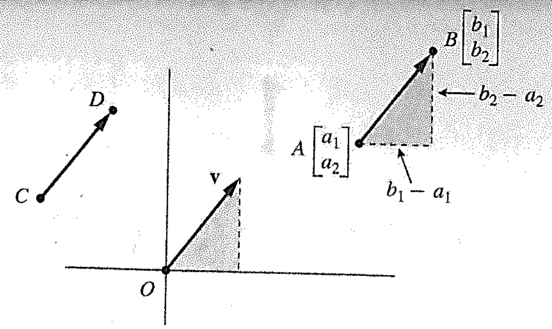


Figure 1.2

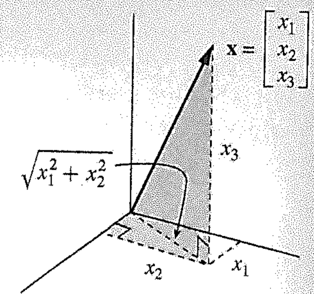


Figure 1.3

to A as the "tail" of the vector $\overrightarrow{AB}$ and B as its "head." If $A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$, then $\overrightarrow{AB}$ is equal to the vector $\mathbf{v} = \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix}$, whose tail is at the origin, as indicated in Figure 1.2.

The Pythagorean Theorem tells us that when $n = 2$ the length of the vector $\mathbf{x}$ is $\sqrt{x_1^2 + x_2^2}$. A repeated application of the Pythagorean Theorem, as indicated in Figure 1.3, leads to the following

Definition We define the *length* of the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \qquad \text{to be} \qquad \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

We say $\mathbf{x}$ is a *unit vector* if it has length 1, i.e., if $\|\mathbf{x}\| = 1$.

There are two crucial algebraic operations one can perform on vectors, both of which have clear geometric interpretations.

Scalar multiplication: If c is a real number and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a vector, then we define $c\mathbf{x}$ to be the vector $\begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$. Note that $c\mathbf{x}$ points in either the same direction as $\mathbf{x}$ or the opposite direction, depending on whether $c > 0$ or $c < 0$, respectively. Thus, multiplication by the real number c simply stretches (or shrinks) the vector by a factor of $|c|$ and reverses

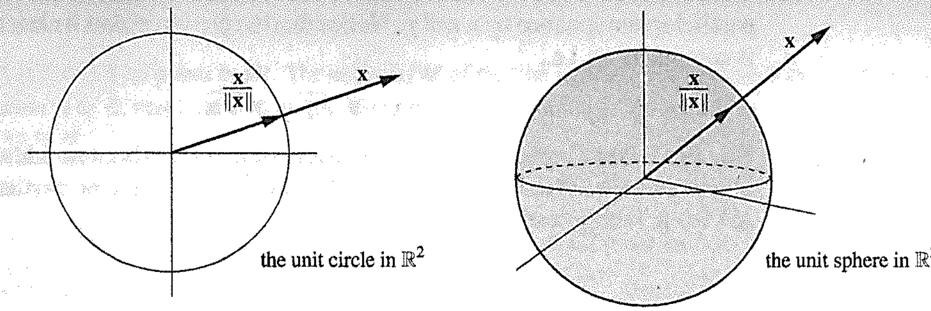


Figure 1.4

its direction when c is negative. Since this is a geometric "change of scale," we refer to the real number c as a *scalar* and the multiplication $c\mathbf{x}$ as *scalar multiplication*.

Note that whenever $\mathbf{x} \neq \mathbf{0}$ we can find a unit vector with the same direction by taking

$$\frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{1}{\|\mathbf{x}\|}\mathbf{x},$$

as shown in Figure 1.4.

Given a nonzero vector $\mathbf{x}$, any scalar multiple $c\mathbf{x}$ lies on the line through the origin and passing through the head of the vector $\mathbf{x}$. For this reason, we make the following

Definition We say two vectors $\mathbf{x}$ and $\mathbf{y}$ are *parallel* if one is a scalar multiple of the other, i.e., if there is a scalar c so that $\mathbf{y} = c\mathbf{x}$ or $\mathbf{x} = c\mathbf{y}$. We say $\mathbf{x}$ and $\mathbf{y}$ are *nonparallel* if they are not parallel.

Vector addition: If $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, then we define $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$.

To understand this geometrically, we move the vector $\mathbf{y}$ so that its tail is at the head of $\mathbf{x}$, and draw the arrow from the origin to its head. This is the so-called *parallelogram law* for vector addition, for, as we see in Figure 1.5, $\mathbf{x} + \mathbf{y}$ is the "long" diagonal of the

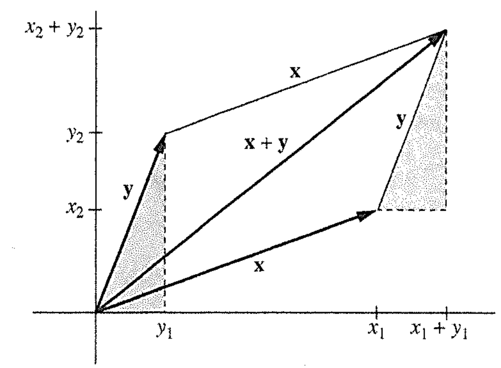


Figure 1.5

parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$. Notice that the picture makes it clear that vector addition is commutative; i.e.,

$$\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}.$$

This also follows immediately from the algebraic definition because addition of real numbers is commutative. (See Exercise 12 for an exhaustive list of the properties of vector addition and scalar multiplication.)

Remark We emphasize here that the notions of vector addition and scalar multiplication make sense geometrically for vectors in the form $\overrightarrow{AB}$ which do not necessarily have their tails at the origin. If we wish to add $\overrightarrow{AB}$ to $\overrightarrow{CD}$, we simply recall that $\overrightarrow{CD}$ is equal to *any* vector with the same length and direction, so we just translate $\overrightarrow{CD}$ so that C and B coincide; then the arrow from A to the point D in its new position is the sum $\overrightarrow{AB}+\overrightarrow{CD}$.

Subtraction of one vector from another is easy to define algebraically. If $\mathbf{x}$ and $\mathbf{y}$ are as above, then we set

$$\mathbf{x}-\mathbf{y}=\begin{bmatrix} x_1-y_1 \\ \vdots \\ x_n-y_n \end{bmatrix}.$$

As is the case with real numbers, we have the following interpretation of the difference $\mathbf{x}-\mathbf{y}$: It is the vector we add to $\mathbf{y}$ in order to obtain $\mathbf{x}$; i.e.,

$$(\mathbf{x}-\mathbf{y})+\mathbf{y}=\mathbf{x}.$$

Pictorially, we see that $\mathbf{x}-\mathbf{y}$ is drawn, as shown in Figure 1.6, by putting its tail at $\mathbf{y}$ and its head at $\mathbf{x}$, thereby resulting in the other diagonal of the parallelogram determined by $\mathbf{x}$ and $\mathbf{y}$. Note that if A and B are points in space and we set $\mathbf{x}=\overrightarrow{OA}$ and $\mathbf{y}=\overrightarrow{OB}$, then $\mathbf{y}-\mathbf{x}=\overrightarrow{AB}$. Moreover, as Figure 1.6 also suggests, we have $\mathbf{x}-\mathbf{y}=\mathbf{x}+(-\mathbf{y})$.

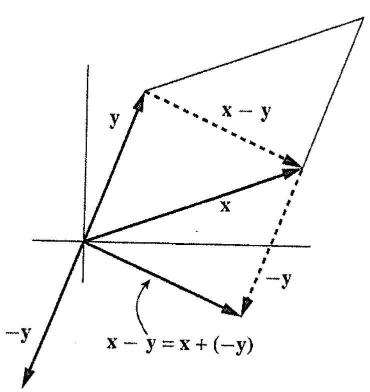


Figure 1.6

EXAMPLE 1

Let A and B be points in $\mathbb{R}^n$. The *midpoint* M of the line segment joining them is the point halfway from A to B; that is, $\overrightarrow{AM}=\frac{1}{2}\overrightarrow{AB}$. Using the notation as above, we set $\mathbf{x}=\overrightarrow{OA}$ and $\mathbf{y}=\overrightarrow{OB}$, and we have

$$(*)\qquad \overrightarrow{OM}=\mathbf{x}+\overrightarrow{AM}=\mathbf{x}+\tfrac{1}{2}(\mathbf{y}-\mathbf{x})=\tfrac{1}{2}(\mathbf{x}+\mathbf{y}).$$

In particular, the vector from the origin to the midpoint of $\overline{AB}$ is the average of the vectors $\mathbf{x}$ and $\mathbf{y}$. See Exercise 8 for a generalization to three vectors and Section 4 of Chapter 7 for more.

From this formula follows one of the classic results from high school geometry: The diagonals of a parallelogram bisect one another. We've seen that the midpoint M of $\overline{AB}$ is, by virtue of the formula (∗), also the midpoint of diagonal $\overline{OC}$. (See Figure 1.7.) ◄

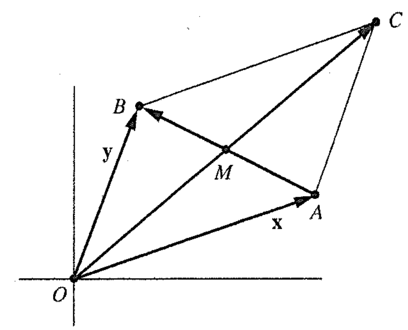


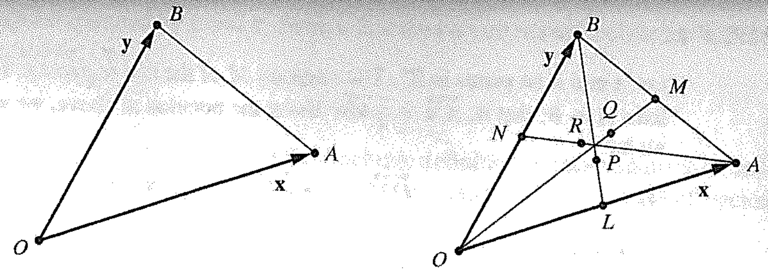
Figure 1.7

It should now be evident that vector methods provide a great tool for translating theorems from Euclidean geometry into simple algebraic statements. Here is another example. Recall that a *median* of a triangle is a line segment from a vertex to the midpoint of the opposite side.

Proposition 1.1 *The medians of a triangle intersect at a point that is two-thirds of the way from each vertex to the opposite side.*

Proof We may put one of the vertices of the triangle at the origin, so that the picture is as shown in Figure 1.8(a). Let $\mathbf{x}=\overrightarrow{OA}$, $\mathbf{y}=\overrightarrow{OB}$, and let L, M, and N be the midpoints of $\overline{OA}$, $\overline{AB}$, and $\overline{OB}$, respectively. The battle plan is the following: We let P denote the point 2/3 of the way from B to L, Q the point 2/3 of the way from O to M, and R the point 2/3 of the way from A to N. Although we've indicated P, Q, and R as distinct points in Figure 1.8(b), our goal is to prove that $P=Q=R$; we do this by expressing all the vectors $\overrightarrow{OP}$, $\overrightarrow{OQ}$, and $\overrightarrow{OR}$ in terms of $\mathbf{x}$ and $\mathbf{y}$.

$$\begin{aligned}\overrightarrow{OP}&=\overrightarrow{OB}+\overrightarrow{BP}=\overrightarrow{OB}+\tfrac{2}{3}\overrightarrow{BL}=\mathbf{y}+\tfrac{2}{3}\left(\tfrac{1}{2}\mathbf{x}-\mathbf{y}\right)\\&=\tfrac{1}{3}\mathbf{x}+\tfrac{1}{3}\mathbf{y};\\ \overrightarrow{OQ}&=\tfrac{2}{3}\overrightarrow{OM}=\tfrac{2}{3}\left(\tfrac{1}{2}(\mathbf{x}+\mathbf{y})\right)=\tfrac{1}{3}(\mathbf{x}+\mathbf{y});\ \text{and}\end{aligned}$$



(a) (b)

Figure 1.8

$$\overrightarrow{OR} = \overrightarrow{OA} + \overrightarrow{AR} = \overrightarrow{OA} + \tfrac{2}{3}\overrightarrow{AN} = \mathbf{x} + \tfrac{2}{3}(\tfrac{1}{2}\mathbf{y} - \mathbf{x}) = \tfrac{1}{3}\mathbf{x} + \tfrac{1}{3}\mathbf{y}.$$

We conclude that, as desired, $\overrightarrow{OP} = \overrightarrow{OQ} = \overrightarrow{OR}$, and so $P = Q = R$. That is, if we go 2/3 of the way down any of the medians, we end up at the same point; this is, of course, the point of intersection of the three medians. ■

The astute reader might notice that we could have been more economical in the last proof. Suppose we merely check that the points 2/3 of the way down *two* of the medians (say P and Q) agree. It would then follow (say, by relabeling the triangle slightly) that the same is true of a different pair of medians (say P and R). But since any two pairs must have a point in common, we may now conclude that all three points are equal.

EXERCISES 1.1

1. Given $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, calculate the following both algebraically and geometrically.

(a) $\mathbf{x} + \mathbf{y}$
(b) $\mathbf{x} - \mathbf{y}$
(c) $\mathbf{x} + 2\mathbf{y}$
(d) $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}$
(e) $\mathbf{y} - \mathbf{x}$
(f) $2\mathbf{x} - \mathbf{y}$
(g) $\|\mathbf{x}\|$
(h) $\mathbf{x}/\|\mathbf{x}\|$

***2.** Three vertices of a parallelogram are $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$, and $\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$. What are all the possible positions of the fourth vertex? Give your reasoning.

3. The origin is at the center of a regular polygon.
(a) What is the sum of the vectors to each of the vertices of the polygon? Give your reasoning. (Hint: What are the symmetries of the polygon?)
(b) What is the sum of the vectors from one fixed vertex to each of the remaining vertices? Give your reasoning.

4. Given $\triangle ABC$, let M and N be the midpoints of $\overline{AB}$ and $\overline{AC}$, respectively. Prove that $\overrightarrow{MN} = \frac{1}{2}\overrightarrow{BC}$.

5. Let $ABCD$ be an arbitrary quadrilateral. Let P, Q, R, and S be the midpoints of $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, and $\overline{DA}$, respectively. Use vector methods to prove that $PQRS$ is a parallelogram. (Hint: Use Exercise 4.)

***6.** In $\triangle ABC$ pictured in Figure 1.9, $\|\overrightarrow{AD}\| = \frac{2}{3}\|\overrightarrow{AB}\|$ and $\|\overrightarrow{CE}\| = \frac{2}{5}\|\overrightarrow{CB}\|$. Let Q denote the midpoint of $\overline{CD}$; show that $\overrightarrow{AQ} = c\overrightarrow{AE}$ for some scalar c and determine the ratio $c = \|\overrightarrow{AQ}\|/\|\overrightarrow{AE}\|$. In what ratio does $\overline{CD}$ divide $\overline{AE}$?

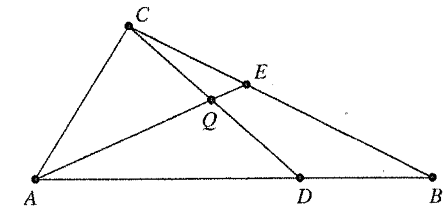


Figure 1.9

7. Consider parallelogram $ABCD$. Suppose $\overrightarrow{AE} = \frac{1}{3}\overrightarrow{AB}$ and $\overrightarrow{DP} = \frac{3}{4}\overrightarrow{DE}$. Show that P lies on the diagonal $\overline{AC}$. (See Figure 1.10.)

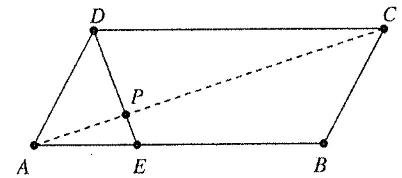


Figure 1.10

8. Let A, B, and C be vertices of a triangle in $\mathbb{R}^3$. Let $\mathbf{x} = \overrightarrow{OA}$, $\mathbf{y} = \overrightarrow{OB}$, and $\mathbf{z} = \overrightarrow{OC}$. Show that the head of the vector $\mathbf{v} = \frac{1}{3}(\mathbf{x} + \mathbf{y} + \mathbf{z})$ lies on each median of $\triangle ABC$ (and thus is the point of intersection of the three medians). It follows (see Section 4 of Chapter 7) that when we put equal masses at A, B, and C, the center of mass of that system is given by the intersections of the medians of the triangle.

9. **(a)** Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. Describe the vectors $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$, where $s + t = 1$. Pay particular attention to the location of $\mathbf{x}$ when $s \geq 0$ and when $t \geq 0$.
(b) Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Describe the vectors $\mathbf{x} = r\mathbf{u} + s\mathbf{v} + t\mathbf{w}$, where $r + s + t = 1$. Pay particular attention to the location of $\mathbf{x}$ when each of r, s, and t is positive.

10. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are nonparallel vectors. (Recall the definition on p. 3.)
(a) Prove that if $s\mathbf{x} + t\mathbf{y} = \mathbf{0}$, then $s = t = 0$. (Hint: Show that neither $s \neq 0$ nor $t \neq 0$ is possible.)
(b) Prove that if $a\mathbf{x} + b\mathbf{y} = c\mathbf{x} + d\mathbf{y}$, then $a = c$ and $b = d$.

11. "Discover" the fraction 2/3 that appears in Proposition 1.1 by finding the intersection of two medians. (Hint: A point on the line $\overleftrightarrow{OM}$ can be written in the form $t(\mathbf{x} + \mathbf{y})$ for some scalar t, and a point on the line $\overleftrightarrow{AN}$ can be written in the form $\mathbf{x} + s(\frac{1}{2}\mathbf{y} - \mathbf{x})$ for some scalar s. You will need to use the result of Exercise 10.)

12. Verify both algebraically and geometrically that the following properties of vector arithmetic hold. (Do so for $n = 2$ if the general case is too intimidating.)
(a) For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
(b) For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
(c) $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

(d) For each $\mathbf{x} \in \mathbb{R}^n$, there is a vector $-\mathbf{x}$ so that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

(e) For all $c, d \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $c(d\mathbf{x}) = (cd)\mathbf{x}$.

(f) For all $c \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $c(\mathbf{x}+\mathbf{y}) = c\mathbf{x} + c\mathbf{y}$.

(g) For all $c, d \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $(c+d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$.

(h) For all $\mathbf{x} \in \mathbb{R}^n$, $1\mathbf{x} = \mathbf{x}$.

*13. (a) Using only the properties listed in Exercise 12, prove that for any $\mathbf{x} \in \mathbb{R}^n$, we have $0\mathbf{x} = \mathbf{0}$. (It often surprises students that this is a consequence of the properties in Exercise 12.)

(b) Using the result of part a, prove that $(-1)\mathbf{x} = -\mathbf{x}$. (Be sure that you didn't use this fact in your proof of part a!)

2 DOT PRODUCT

We discuss next one of the crucial constructions in linear algebra, the dot product $\mathbf{x} \cdot \mathbf{y}$ of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By way of motivation, let's recall some basic results from plane geometry. Let $P = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $Q = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ be points in the plane, as pictured in Figure 2.1.

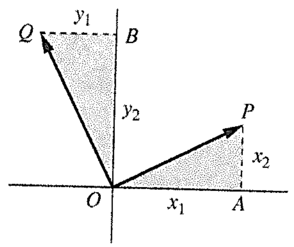


Figure 2.1

Then we observe that when $\angle POQ$ is a right angle, ΔOAP is similar to ΔOBQ, and so $x_2/x_1 = -y_1/y_2$, whence $x_1y_1 + x_2y_2 = 0$. This leads us to make the following

Definition Given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, define their *dot product*

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2.$$

More generally, given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, define their dot product

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

We know that when the vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^2$ are perpendicular, their dot product is 0. By starting with the algebraic properties of the dot product, we are able to get a great deal of geometry out of it.

Proposition 2.1 *The dot product has the following properties:*

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *(dot product is commutative);*
2. $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2 \geq 0$ *and* $\mathbf{x} \cdot \mathbf{x} = 0 \iff \mathbf{x} = \mathbf{0}$*;*
3. $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$*;*
4. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$ *for all* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ *(the distributive property).*

Proof In order to simplify the notation, we give the proof with $n = 2$. Since multiplication of real numbers is commutative, we have

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 = y_1x_1 + y_2x_2 = \mathbf{y} \cdot \mathbf{x}.$$

The square of a real number is nonnegative and the sum of nonnegative numbers is nonnegative, so $\mathbf{x} \cdot \mathbf{x} = x_1^2 + x_2^2 \geq 0$ and is equal to 0 only when $x_1 = x_2 = 0$. The next property follows from the associative and distributive properties of real numbers:

$$(c\mathbf{x}) \cdot \mathbf{y} = (cx_1)y_1 + (cx_2)y_2 = c(x_1y_1) + c(x_2y_2) = c(x_1y_1 + x_2y_2) = c(\mathbf{x} \cdot \mathbf{y}).$$

The last result follows from the commutative, associative, and distributive properties of real numbers:

$$\begin{aligned} \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= x_1(y_1 + z_1) + x_2(y_2 + z_2) = x_1y_1 + x_1z_1 + x_2y_2 + x_2z_2 \\ &= (x_1y_1 + x_2y_2) + (x_1z_1 + x_2z_2) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}. \ \blacksquare \end{aligned}$$

Corollary 2.2 $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2$.

Proof Using the properties repeatedly, we have

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} \\ &= \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2, \end{aligned}$$

as desired. $\blacksquare$

The geometric meaning of this result comes from the Pythagorean Theorem: When $\mathbf{x}$ and $\mathbf{y}$ are perpendicular vectors in $\mathbb{R}^2$, then we have $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, and so, by Corollary 2.2, it must be the case that $\mathbf{x} \cdot \mathbf{y} = 0$. (And the converse follows, too, from the converse of the Pythagorean Theorem.) That is, two vectors in $\mathbb{R}^2$ are perpendicular if and only if their dot product is 0.

Motivated by this, we use the algebraic definition of dot product of vectors in $\mathbb{R}^n$ to bring in the geometry. In keeping with current use of the terminology and falling prey to the penchant to have several names for the same thing, we make the following

Definition We say vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x} \cdot \mathbf{y} = 0$.

Armed with this definition, we proceed to a construction that will be important in much of our future work. Starting with two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{y} \neq \mathbf{0}$, Figure 2.2 suggests that we should be able to write $\mathbf{x}$ as the sum of a vector, $\mathbf{x}^{\parallel}$, that is parallel to $\mathbf{y}$ and a vector, $\mathbf{x}^{\perp}$, that is orthogonal to $\mathbf{y}$. Let's suppose we have such an equation:

$$\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}, \quad \text{where}$$

$$\mathbf{x}^{\parallel} \text{ is a scalar multiple of } \mathbf{y} \quad \text{and} \quad \mathbf{x}^{\perp} \text{ is orthogonal to } \mathbf{y}.$$

To say that $\mathbf{x}^{\parallel}$ is a scalar multiple of $\mathbf{y}$ means that we can write $\mathbf{x}^{\parallel} = c\mathbf{y}$ for some scalar c. Now, assuming such an expression exists, we can determine c by taking the dot product of both sides of the equation with $\mathbf{y}$:

$$\mathbf{x} \cdot \mathbf{y} = (\mathbf{x}^{\parallel} + \mathbf{x}^{\perp}) \cdot \mathbf{y} = (\mathbf{x}^{\parallel} \cdot \mathbf{y}) + (\mathbf{x}^{\perp} \cdot \mathbf{y}) = \mathbf{x}^{\parallel} \cdot \mathbf{y} = (c\mathbf{y}) \cdot \mathbf{y} = c\|\mathbf{y}\|^2.$$

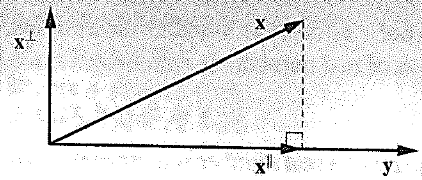


Figure 2.2

This means that

$$c = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}, \quad \text{and so} \quad \mathbf{x}^\parallel = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}.$$

The vector $\mathbf{x}^\parallel$ is called the *projection of* $\mathbf{x}$ *onto* $\mathbf{y}$, written $\text{proj}_\mathbf{y}\mathbf{x}$.

The fastidious reader may be puzzled by the logic here. We have apparently assumed that we can write $\mathbf{x} = \mathbf{x}^\parallel + \mathbf{x}^\perp$ in order to prove that we can do so. Of course, as it stands, this is not fair. Here's how we fix it. We now *define*

$$\mathbf{x}^\parallel = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}$$
$$\mathbf{x}^\perp = \mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}.$$

Obviously, $\mathbf{x}^\parallel + \mathbf{x}^\perp = \mathbf{x}$ and $\mathbf{x}^\parallel$ is a scalar multiple of $\mathbf{y}$. All we need to check is that $\mathbf{x}^\perp$ is in fact orthogonal to $\mathbf{y}$. Well,

$$\mathbf{x}^\perp\cdot\mathbf{y} = \left(\mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}\right)\cdot\mathbf{y} = \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}\cdot\mathbf{y} = \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\|\mathbf{y}\|^2 = \mathbf{x}\cdot\mathbf{y} - \mathbf{x}\cdot\mathbf{y} = 0,$$

as required. Note, moreover, that $\mathbf{x}^\parallel$ is the *unique* multiple of $\mathbf{y}$ that satisfies the equation $(\mathbf{x} - \mathbf{x}^\parallel)\cdot\mathbf{y} = 0$.

EXAMPLE 1

Let $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$. Then

$$\mathbf{x}^\parallel = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} = \frac{\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}}{\left\|\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}\right\|^2}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{x}^\perp = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{8}{3} \\ \frac{7}{3} \\ \frac{1}{3} \end{bmatrix}.$$

To double-check, we compute $\mathbf{x}^\perp\cdot\mathbf{y} = \begin{bmatrix} \frac{8}{3} \\ \frac{7}{3} \\ \frac{1}{3} \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = 0$, as it should be. ◀

Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. We shall see next that the formula for the projection of $\mathbf{x}$ onto $\mathbf{y}$ enables us to calculate the *angle* between the vectors $\mathbf{x}$ and $\mathbf{y}$. Consider the right triangle in Figure 2.3; let θ denote the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. Remembering that the cosine of an angle is the ratio of the *signed* length of the adjacent side to the length of the hypotenuse, we see that

$$\cos\theta = \frac{\text{signed length of } \mathbf{x}^\parallel}{\text{length of } \mathbf{x}} = \frac{c\|\mathbf{y}\|}{\|\mathbf{x}\|} = \frac{\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\|\mathbf{y}\|}{\|\mathbf{x}\|} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

This, then, is the geometric interpretation of the dot product:

$$\boxed{\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.}$$

Will this formula still make sense even when $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$? Geometrically, we simply restrict our attention to the plane spanned by $\mathbf{x}$ and $\mathbf{y}$ and measure the angle θ in that plane, and so we blithely make the

Definition Let $\mathbf{x}$ and $\mathbf{y}$ be nonzero vectors in $\mathbb{R}^n$. We define the *angle* between them to be the unique θ satisfying $0 \le \theta \le \pi$ so that

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

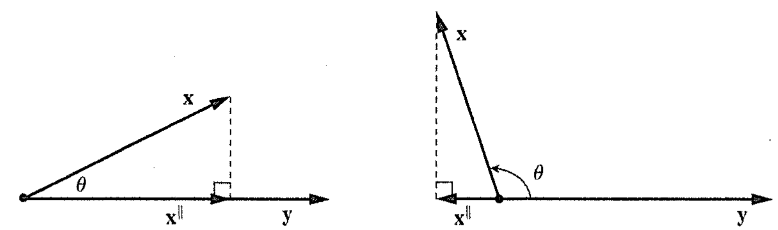


Figure 2.3

Since our geometric intuition may be misleading in $\mathbb{R}^n$, we should check *algebraically* that this definition makes sense. Since $|\cos\theta| \le 1$, the following result gives us what is needed.

Proposition 2.3 (Cauchy-Schwarz Inequality) *If* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *then*

$$|\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\|.$$

Moreover, equality holds if and only if one of the vectors is a scalar multiple of the other.

Proof If $\mathbf{y} = \mathbf{0}$, then there's nothing to prove. If $\mathbf{y} \neq \mathbf{0}$, then we observe that the quadratic function of t given by

$$g(t) = \|\mathbf{x} + t\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2t\mathbf{x} \cdot \mathbf{y} + t^2\|\mathbf{y}\|^2$$

takes its minimum at $t_0 = -\dfrac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{y}\|^2}$. The minimum value

$$g(t_0) = \|\mathbf{x}\|^2 - 2\frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} + \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2}$$

is necessarily nonnegative, so

$$(\mathbf{x} \cdot \mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2,$$

and, since square root preserves inequality,

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|,$$

as desired. Equality holds if and only if $\mathbf{x} + t\mathbf{y} = \mathbf{0}$ for some scalar t. (See Exercise 9 for a discussion of how this proof relates to our formula for $\text{proj}_{\mathbf{y}}\mathbf{x}$ above.) ■

One of the most useful applications of this result is the famed *triangle inequality*, which tells us that the sum of the lengths of two sides of a triangle cannot be less than the length of the third.

Corollary 2.4 (Triangle Inequality) *For any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.*

Proof By Corollary 2.2 and Proposition 2.3 we have

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.$$

Since square root preserves inequality, we conclude that $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, as desired. ■

Remark The dot product also arises in situations removed from geometry. The economist introduces the *commodity vector* $\mathbf{x}$, whose entries are the quantities of various commodities that happen to be of interest and the *price vector* $\mathbf{p}$. For example, we might consider

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \quad \text{and} \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} \in \mathbb{R}^5,$$

where x_1 represents the number of pounds of flour, x_2 the number of dozens of eggs, x_3 the number of pounds of chocolate chips, x_4 the number of pounds of walnuts, and x_5 the number of pounds of butter needed to produce a certain massive quantity of chocolate chip cookies, and p_i is the price (in dollars) of a unit of the i^{th} commodity (e.g., p_2 is the price of a dozen eggs). Then it is easy to see that

$$\mathbf{p} \cdot \mathbf{x} = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 + p_5x_5$$

is the total cost of producing the massive quantity of cookies. (To be realistic, we might also want to include x_6 as the number of hours of labor, with corresponding hourly wage p_6.) We will return to this interpretation in Section 4.

EXERCISES 1.2

1. For each of the following pairs of vectors $\mathbf{x}$ and $\mathbf{y}$, calculate $\mathbf{x} \cdot \mathbf{y}$ and the angle θ between the vectors.

(a) $\mathbf{x} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -5 \\ 2 \end{bmatrix}$

(b) $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

*(c) $\mathbf{x} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

(d) $\mathbf{x} = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix}$

(e) $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 6 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}$

*(f) $\mathbf{x} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

(g) $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ -3 \\ -1 \\ 5 \end{bmatrix}$

*2. For each pair of vectors in Exercise 1, calculate $\text{proj}_{\mathbf{y}}\mathbf{x}$ and $\text{proj}_{\mathbf{x}}\mathbf{y}$.

*3. Find the angle between the long diagonal of a cube and a face diagonal.

4. Find the angle that the long diagonal of a $3 \times 4 \times 5$ rectangular box makes with the longest edge.

5. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x}\| = 2$, $\|\mathbf{y}\| = 1$, and the angle θ between $\mathbf{x}$ and $\mathbf{y}$ is $\theta = \arccos(1/4)$. Prove that the vectors $\mathbf{x} - 3\mathbf{y}$ and $\mathbf{x} + \mathbf{y}$ are orthogonal.

6. Suppose $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ are unit vectors satisfying $\mathbf{x} + \mathbf{y} + \mathbf{z} = \mathbf{0}$. What can you say about the angles between each pair?

7. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ be the so-called *standard basis vectors* for $\mathbb{R}^3$. Let $\mathbf{x} \in \mathbb{R}^3$ be a nonzero vector. For $i = 1, 2, 3$, let θ_i denote the angle between $\mathbf{x}$ and $\mathbf{e}_i$. Compute $\cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3$.

*8. Let $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{bmatrix} \in \mathbb{R}^n$. Let θ_n be the angle between $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. Find $\lim_{n\to\infty} \theta_n$.

(Hint: You may need to recall the formulas for $1 + 2 + \cdots + n$ and $1^2 + 2^2 + \cdots + n^2$ from your beginning calculus course.)

9. With regard to the proof of Proposition 2.3, how is $t_0\mathbf{y}$ related to $\mathbf{x}^{\parallel}$? What does this say about $\text{proj}_{\mathbf{y}}\mathbf{x}$?

10. Use vector methods to prove that a parallelogram is a rectangle if and only if its diagonals have the same length.

11. Use the fundamental properties of the dot product to prove that

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right).$$

Interpret the result geometrically.

***12.** Use the dot product to prove the law of cosines: As shown in Figure 2.4,

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

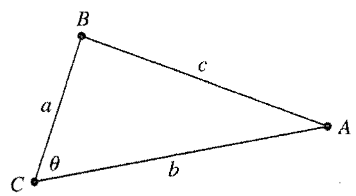


Figure 2.4

13. Use vector methods to prove that the diagonals of a parallelogram are orthogonal if and only if the parallelogram is a rhombus (i.e., has all sides of equal length).

‡14. Use vector methods to prove that a triangle inscribed in a circle and having a diameter as one of its sides must be a right triangle. (Hint: See Figure 2.5.)

Geometric challenge: More generally, given two points A and B in the plane, what is the locus of points X so that $\angle AXB$ has a fixed measure?

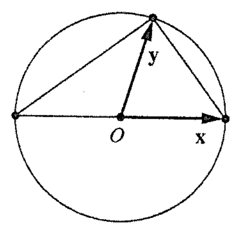


Figure 2.5

15. (a) Let $\mathbf{y} \in \mathbb{R}^n$. If $\mathbf{x}\cdot\mathbf{y} = 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$, then prove that $\mathbf{y} = \mathbf{0}$.

(b) Suppose $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot\mathbf{z}$ for all $\mathbf{x} \in \mathbb{R}^n$. What can you conclude?

16. If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, set $\rho(\mathbf{x}) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$.

(a) Check that $\rho(\mathbf{x})$ is orthogonal to $\mathbf{x}$; indeed, $\rho(\mathbf{x})$ is obtained by rotating $\mathbf{x}$ an angle $\pi/2$ counterclockwise.

(b) Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, prove that $\mathbf{x}\cdot\rho(\mathbf{y}) = -\rho(\mathbf{x})\cdot\mathbf{y}$. Interpret this statement geometrically.

‡17. Prove that for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have $\|\mathbf{x}\| - \|\mathbf{y}\| \le \|\mathbf{x}-\mathbf{y}\|$. Deduce that $\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \le \|\mathbf{x}-\mathbf{y}\|$. (Hint: Apply the result of Corollary 2.4 directly.)

18. Use the Cauchy-Schwarz inequality to solve the following max/min problem: If the (long) diagonal of a rectangular box has length c, what is the greatest the sum of the length, width, and height of the box can be? For what shape box does the maximum occur?

19. Give an alternative proof of the Cauchy-Schwarz inequality, as follows. Let $a = \|\mathbf{x}\|$, $b = \|\mathbf{y}\|$, and deduce from $\|b\mathbf{x} - a\mathbf{y}\|^2 \ge 0$ that $\mathbf{x}\cdot\mathbf{y} \le ab$. Now how do you show that $|\mathbf{x}\cdot\mathbf{y}| \le ab$? When does equality hold?

‡20. (a) Let $\mathbf{x}$ and $\mathbf{y}$ be vectors with $\|\mathbf{x}\| = \|\mathbf{y}\|$. Prove that the vector $\mathbf{x}+\mathbf{y}$ bisects the angle between $\mathbf{x}$ and $\mathbf{y}$.

(b) More generally, if $\mathbf{x}$ and $\mathbf{y}$ are arbitrary nonzero vectors, let $a = \|\mathbf{x}\|$ and $b = \|\mathbf{y}\|$. Prove that the vector $b\mathbf{x} + a\mathbf{y}$ bisects the angle between $\mathbf{x}$ and $\mathbf{y}$.

21. Use vector methods to prove that the diagonals of a parallelogram bisect the vertex angles if and only if the parallelogram is a rhombus.

22. Given $\triangle ABC$ with D on $\overline{BC}$ as shown in Figure 2.6. Prove that if $\overrightarrow{AD}$ bisects $\angle BAC$, then $\|\overrightarrow{BD}\|/\|\overrightarrow{CD}\| = \|\overrightarrow{AB}\|/\|\overrightarrow{AC}\|$. (Hint: Use Exercise 20b. Let $\mathbf{x} = \overrightarrow{AB}$ and $\mathbf{y} = \overrightarrow{AC}$; give two expressions for $\overrightarrow{AD}$ in terms of $\mathbf{x}$ and $\mathbf{y}$ and use Exercise 1.1.10.)

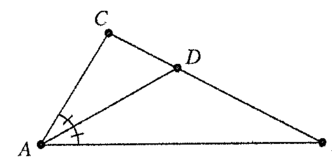


Figure 2.6

23. Use vector methods to prove that the angle bisectors of a triangle have a common point. (Hint: Given $\triangle OAB$, let $\mathbf{x} = \overrightarrow{OA}$, $\mathbf{y} = \overrightarrow{OB}$, $a = \|\overrightarrow{OA}\|$, $b = \|\overrightarrow{OB}\|$, and $c = \|\overrightarrow{AB}\|$. If we define the point P by $\overrightarrow{OP} = \frac{1}{a+b+c}(b\mathbf{x} + a\mathbf{y})$, use Exercise 20b to show that P lies on all three angle bisectors.)

24. Use vector methods to prove that the altitudes of a triangle have a common point. Recall that altitudes of a triangle are the lines passing through a vertex and perpendicular to the opposite side. (Hint: See Figure 2.7. Let C be the point of intersection of the altitude from B to $\overleftrightarrow{OA}$ and the altitude from A to $\overleftrightarrow{OB}$. Prove that $\overrightarrow{OC}$ is orthogonal to $\overrightarrow{AB}$.)

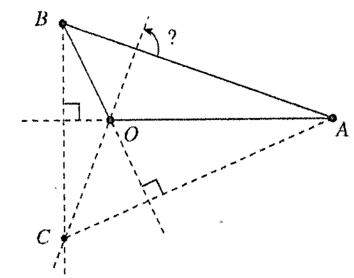


Figure 2.7

25. Use vector methods to prove that the perpendicular bisectors of the sides of a triangle intersect in a point, as follows. Assume the triangle OAB has one vertex at the origin, and let $\mathbf{x} = \overrightarrow{OA}$ and $\mathbf{y} = \overrightarrow{OB}$.

(a) Let $\mathbf{z}$ be the point of intersection of the perpendicular bisectors of $\overline{OA}$ and $\overline{OB}$. Prove that (using the notation of Exercise 16)

$$\mathbf{z} = \tfrac{1}{2}\mathbf{x} + c\rho(\mathbf{x}), \quad \text{where} \quad c = \frac{\|\mathbf{y}\|^2 - \mathbf{x}\cdot\mathbf{y}}{2\rho(\mathbf{x})\cdot\mathbf{y}}.$$

(b) Show that $\mathbf{z}$ lies on the perpendicular bisector of $\overline{AB}$. (Hint: What is the dot product of $\mathbf{z} - \frac{1}{2}(\mathbf{x}+\mathbf{y})$ with $\mathbf{y}-\mathbf{x}$?)

26. Let P be the intersection of the medians of $\triangle OAB$ (see Proposition 1.1), Q the intersection of its altitudes (see Exercise 24), and R the intersection of the perpendicular bisectors of its sides (see Exercise 25). Show that P, Q, and R are collinear and that P is two-thirds of the way from Q to R. Does the intersection of the angle bisectors (see Exercise 23) lie on this line as well?

3 SUBSPACES OF $\mathbb{R}^n$

As we proceed in our study of "linear objects," it is fundamental to concentrate on subsets of $\mathbb{R}^n$ that are generalizations of lines and planes through the origin.

Definition A set $V \subset \mathbb{R}^n$ (a *subset* of $\mathbb{R}^n$) is called a *subspace* of $\mathbb{R}^n$ if it satisfies the following properties:

1. $\mathbf{0} \in V$ (the zero vector belongs to V);
2. whenever $\mathbf{v} \in V$ and $c \in \mathbb{R}$, we have $c\mathbf{v} \in V$ (V is closed under scalar multiplication);
3. whenever $\mathbf{v}, \mathbf{w} \in V$, we have $\mathbf{v}+\mathbf{w} \in V$ (V is closed under addition).

EXAMPLE 1

Let's begin with some familiar examples.

a. The *trivial subspace* consisting of just the zero vector $\mathbf{0} \in \mathbb{R}^n$ is a subspace since $c\mathbf{0} = \mathbf{0}$ for any scalar c and $\mathbf{0}+\mathbf{0} = \mathbf{0}$.

b. $\mathbb{R}^n$ itself is likewise a subspace of $\mathbb{R}^n$.

c. Fix a nonzero vector $\mathbf{u} \in \mathbb{R}^n$, and consider

$$\ell = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = t\mathbf{u} \text{ for some } t \in \mathbb{R}\}.$$

We check that the three criteria hold:

1. Setting $t = 0$, we see that $\mathbf{0} \in \ell$.
2. If $\mathbf{v} \in \ell$ and $c \in \mathbb{R}$, then $\mathbf{v} = t\mathbf{u}$ for some $t \in \mathbb{R}$, and so $c\mathbf{v} = c(t\mathbf{u}) = (ct)\mathbf{u}$, which is again a scalar multiple of $\mathbf{u}$ and hence an element of ℓ.
3. If $\mathbf{v}, \mathbf{w} \in \ell$, this means that $\mathbf{v} = s\mathbf{u}$ and $\mathbf{w} = t\mathbf{u}$ for some scalars s and t. Then $\mathbf{v}+\mathbf{w} = s\mathbf{u}+t\mathbf{u} = (s+t)\mathbf{u}$, so $\mathbf{v}+\mathbf{w} \in \ell$, as needed.

ℓ is called a *line* through the origin.

d. Fix two nonparallel vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^n$. Set

$$\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = s\mathbf{u}+t\mathbf{v} \text{ for some } s, t \in \mathbb{R}\},$$

as shown in Figure 3.1. $\mathcal{P}$ is called a *plane* through the origin. To see that $\mathcal{P}$ is a subspace, we do the obligatory checks:

1. Setting s and $t = 0$, we see that $\mathbf{0} = 0\mathbf{u}+0\mathbf{v}$, so $\mathbf{0} \in \mathcal{P}$.
2. Suppose $\mathbf{x} \in \mathcal{P}$ and $c \in \mathbb{R}$. Then $\mathbf{x} = s\mathbf{u}+t\mathbf{v}$ for some scalars s and t, and $c\mathbf{x} = c(s\mathbf{u}+t\mathbf{v}) = (cs)\mathbf{u}+(ct)\mathbf{v}$, so $c\mathbf{x} \in \mathcal{P}$ as well.
3. Suppose $\mathbf{x}, \mathbf{y} \in \mathcal{P}$. This means that $\mathbf{x} = s\mathbf{u}+t\mathbf{v}$ for some scalars s and t, and $\mathbf{y} = s'\mathbf{u}+t'\mathbf{v}$ for some scalars s' and t'. Then

$$\mathbf{x}+\mathbf{y} = (s\mathbf{u}+t\mathbf{v})+(s'\mathbf{u}+t'\mathbf{v}) = (s+s')\mathbf{u}+(t+t')\mathbf{v},$$

so $\mathbf{x}+\mathbf{y} \in \mathcal{P}$, as required.

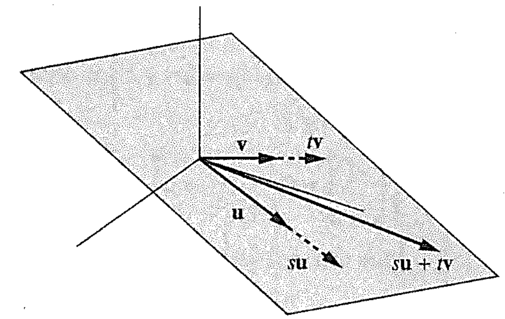


Figure 3.1

e. Fix a nonzero vector $\mathbf{A} \in \mathbb{R}^n$, and consider

$$V = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\cdot\mathbf{x} = 0\}.$$

V consists of all vectors orthogonal to the given vector $\mathbf{A}$, as pictured in Figure 3.2. We check once again that the three criteria hold:

1. Since $\mathbf{A}\cdot\mathbf{0} = 0$, we know that $\mathbf{0} \in V$.
2. Suppose $\mathbf{v} \in V$ and $c \in \mathbb{R}$. Then $\mathbf{A}\cdot(c\mathbf{v}) = c(\mathbf{A}\cdot\mathbf{v}) = 0$, so $c\mathbf{v} \in V$.

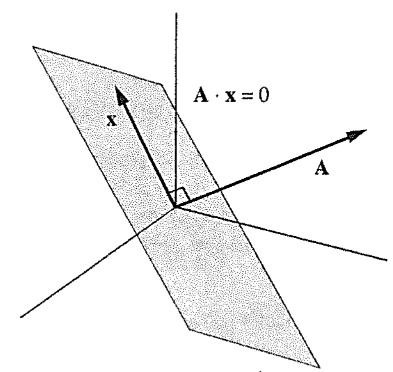


Figure 3.2

3. Suppose $\mathbf{v}, \mathbf{w} \in V$. Then $\mathbf{A} \cdot (\mathbf{v}+\mathbf{w}) = (\mathbf{A}\cdot\mathbf{v}) + (\mathbf{A}\cdot\mathbf{w}) = 0+0 = 0$, so $\mathbf{v}+\mathbf{w} \in V$, as required.

Thus, V is a subspace of $\mathbb{R}^n$. We call V a *hyperplane* in $\mathbb{R}^n$, having *normal vector* $\mathbf{A}$. More generally, given any collection of vectors $\mathbf{A}_1, \dots, \mathbf{A}_m \in \mathbb{R}^n$, the set of solutions of the homogeneous system of linear equations

$$\mathbf{A}_1 \cdot \mathbf{x} = 0, \quad \mathbf{A}_2 \cdot \mathbf{x} = 0, \quad \dots, \quad \mathbf{A}_m \cdot \mathbf{x} = 0$$

forms a subspace of $\mathbb{R}^n$. ◀

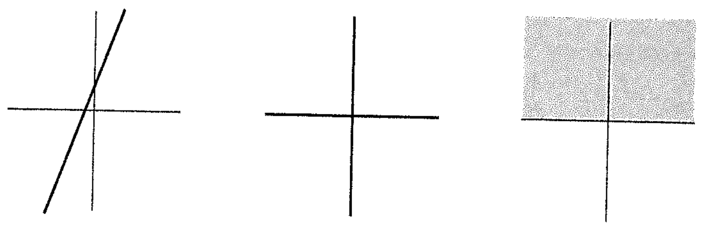
EXAMPLE 2

Let's consider next a few subsets of $\mathbb{R}^2$, as pictured in Figure 3.3, that are *not* subspaces.

a. $S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_2 = 2x_1 + 1 \right\}$ is not a subspace. All three criteria fail, but it suffices to point out $\mathbf{0} \notin S$.

b. $S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_1 x_2 = 0 \right\}$ is not a subspace. Each of the vectors $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ lies in S, and yet their sum $\mathbf{v}+\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ does not.

c. $S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_2 \geq 0 \right\}$ is not a subspace. The vector $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ lies in S, and yet any negative scalar multiple of it, e.g., $(-2)\mathbf{v} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$, does not. ◀



(a) (b) (c)

not subspaces of $\mathbb{R}^2$

Figure 3.3

Given a collection of vectors in $\mathbb{R}^n$, it is natural to try to "build" a subspace from them. We begin with some crucial definitions.

Definition Let $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$. If $c_1, \dots, c_k \in \mathbb{R}$, the vector

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

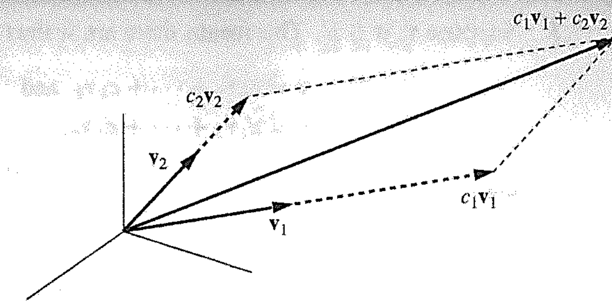


Figure 3.4

(as illustrated in Figure 3.4) is called a *linear combination* of $\mathbf{v}_1, \dots, \mathbf{v}_k$. The set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is called their *span*, denoted $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$.

Every vector in $\mathbb{R}^n$ can be written as a linear combination of the vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are often called the *standard basis* vectors for $\mathbb{R}^n$. Obviously, given the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad \text{we have} \quad \mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Proposition 3.1 *Let* $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$. *Then* $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ *is a subspace of* $\mathbb{R}^n$.

Proof We check that all three criteria hold.

1. To see that $\mathbf{0} \in V$, we merely take $c_1 = c_2 = \cdots = c_k = 0$. Then (by Exercise 1.1.13) $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$.
2. Suppose $\mathbf{v} \in V$ and $c \in \mathbb{R}$. By definition, there are scalars $c_1, \dots, c_k$ so that $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$. Thus,

$$c\mathbf{v} = c(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = (cc_1)\mathbf{v}_1 + (cc_2)\mathbf{v}_2 + \cdots + (cc_k)\mathbf{v}_k,$$

which is again a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$, so $c\mathbf{v} \in V$, as desired.

3. Suppose $\mathbf{v}, \mathbf{w} \in V$. This means there are scalars $c_1, \ldots, c_k$ and $d_1, \ldots, d_k$ so that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \quad \text{and}$$
$$\mathbf{w} = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k;$$

adding, we obtain

$$\begin{aligned}\mathbf{v} + \mathbf{w} &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) + (d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k) \\ &= (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_k + d_k)\mathbf{v}_k,\end{aligned}$$

which is again a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, hence an element of V.

This completes the verification that V is a subspace of $\mathbb{R}^n$. ■

Remark Let $V \subset \mathbb{R}^n$ be a subspace and let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$. We say that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *span* V if $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = V$. (The point here is that every vector in V must be a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$.) As we shall see in Chapter 4, it takes at least n vectors to span $\mathbb{R}^n$; the smallest number of vectors required to span a given subspace will be a measure of its "size" or "dimension."

EXAMPLE 3

The plane

$$\mathcal{P}_1 = \left\{ s\begin{bmatrix}1\\-1\\2\end{bmatrix} + t\begin{bmatrix}2\\0\\1\end{bmatrix} : s, t \in \mathbb{R} \right\}$$

is the span of the vectors

$$\mathbf{v}_1 = \begin{bmatrix}1\\-1\\2\end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix}2\\0\\1\end{bmatrix}$$

and is therefore a subspace of $\mathbb{R}^3$. On the other hand, the plane

$$\mathcal{P}_2 = \left\{ \begin{bmatrix}1\\0\\0\end{bmatrix} + s\begin{bmatrix}1\\-1\\2\end{bmatrix} + t\begin{bmatrix}2\\0\\1\end{bmatrix} : s, t \in \mathbb{R} \right\}$$

is not a subspace. This is most easily verified by checking that $\mathbf{0} \notin \mathcal{P}_2$, for $\mathbf{0} \in \mathcal{P}_2$ precisely when we can find values of s and t so that

$$\begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix} + s\begin{bmatrix}1\\-1\\2\end{bmatrix} + t\begin{bmatrix}2\\0\\1\end{bmatrix}.$$

This amounts to the system of equations:

$$\begin{aligned} s + 2t &= -1 \\ -s \quad &= 0 \\ 2s + t &= 0, \end{aligned}$$

which we easily see has no solution.

A word of warning here: We might have expressed $\mathcal{P}_1$ in the form

$$\left\{ \begin{bmatrix}1\\1\\-1\end{bmatrix} + s\begin{bmatrix}1\\-1\\2\end{bmatrix} + t\begin{bmatrix}2\\0\\1\end{bmatrix} : s, t \in \mathbb{R} \right\},$$

so that, despite the presence of the "shifting" term, the plane may still pass through the origin. ◀

There are really two different ways in which subspaces of $\mathbb{R}^n$ arise: as being the span of a collection of vectors (the "parametric" approach) or as being the set of solutions of a (homogeneous) system of linear equations (the "implicit" approach). We shall study the connections between the two in detail in Chapter 4.

EXAMPLE 4

As the reader can verify, the vector $\mathbf{A} = \begin{bmatrix}-1\\3\\2\end{bmatrix}$ is orthogonal to both the vectors that span the plane $\mathcal{P}_1$ given in Example 3 above. Thus, every vector in $\mathcal{P}_1$ is orthogonal to $\mathbf{A}$, and we suspect that

$$\mathcal{P}_1 = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{A} \cdot \mathbf{x} = 0\} = \{\mathbf{x} \in \mathbb{R}^3 : -x_1 + 3x_2 + 2x_3 = 0\}.$$

Strictly speaking, we only know that every vector in $\mathcal{P}_1$ is a solution of this equation. But note that if $\mathbf{x}$ is a solution, then

$$\mathbf{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}3x_2+2x_3\\x_2\\x_3\end{bmatrix} = x_2\begin{bmatrix}3\\1\\0\end{bmatrix} + x_3\begin{bmatrix}2\\0\\1\end{bmatrix} = (-x_2)\begin{bmatrix}1\\-1\\2\end{bmatrix} + (2x_2 + x_3)\begin{bmatrix}2\\0\\1\end{bmatrix},$$

so $\mathbf{x} \in \mathcal{P}_1$ and the two sets are equal.[1] Thus, the discussion of Example 1e gives another justification that $\mathcal{P}_1$ is a subspace of $\mathbb{R}^3$.

On the other hand, one can check, analogously, that

$$\mathcal{P}_2 = \{\mathbf{x} \in \mathbb{R}^3 : -x_1 + 3x_2 + 2x_3 = -1\},$$

and so clearly $\mathbf{0} \notin \mathcal{P}_2$ and $\mathcal{P}_2$ is not a subspace. It is an *affine* plane parallel to $\mathcal{P}_1$. ◀

Definition Let V and W be subspaces of $\mathbb{R}^n$. We say they are *orthogonal subspaces* if every element of V is orthogonal to every element of W, i.e., if

$$\mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for every } \mathbf{v} \in V \text{ and every } \mathbf{w} \in W.$$

[1] Ordinarily, the easiest way to establish that two sets are equal is to show that each is a subset of the other.

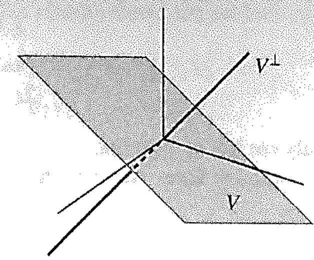


Figure 3.5

As indicated in Figure 3.5, given a subspace $V \subset \mathbb{R}^n$, define

$$V^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{v} = 0 \text{ for every } \mathbf{v} \in V\}.$$

$V^\perp$ (read "V perp") is called the *orthogonal complement* of V.[2]

Proposition 3.2 *$V^\perp$ is also a subspace of $\mathbb{R}^n$.*

Proof We leave this to the reader in Exercise 4. ■

EXAMPLE 5

Let $V = \mathrm{Span}\left(\begin{bmatrix}1\\2\\1\end{bmatrix}\right)$. Then $V^\perp$ is the plane $W = \{x \in \mathbb{R}^3 : x_1 + 2x_2 + x_3 = 0\}$. Now what is the orthogonal complement of W? We suspect it is just the line V, but we will have to wait until Chapter 4 to have the appropriate tools. ◁

If V and W are orthogonal subspaces of $\mathbb{R}^n$, then certainly $W \subset V^\perp$ (why?). Of course, W need not be equal to $V^\perp$: Consider, for example, the x_1-axis and the x_2-axis in $\mathbb{R}^3$.

EXERCISES 1.3

*1. Which of the following are subspaces? Justify your answer in each case.

(a) $\{\mathbf{x} \in \mathbb{R}^2 : x_1 + x_2 = 1\}$

(b) $\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \begin{bmatrix}a\\b\\a+b\end{bmatrix} \text{ for some } a, b \in \mathbb{R}\right\}$

(c) $\{\mathbf{x} \in \mathbb{R}^3 : x_1 + 2x_2 < 0\}$

(d) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$

(e) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 0\}$

(f) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = -1\}$

(g) $\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = s\begin{bmatrix}2\\1\\1\end{bmatrix} + t\begin{bmatrix}1\\2\\1\end{bmatrix} \text{ for some } s, t \in \mathbb{R}\right\}$

(h) $\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \begin{bmatrix}3\\0\\1\end{bmatrix} + s\begin{bmatrix}2\\1\\1\end{bmatrix} + t\begin{bmatrix}1\\2\\1\end{bmatrix} \text{ for some } s, t \in \mathbb{R}\right\}$

(i) $\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \begin{bmatrix}2\\4\\-1\end{bmatrix} + s\begin{bmatrix}2\\1\\1\end{bmatrix} + t\begin{bmatrix}1\\2\\-1\end{bmatrix} \text{ for some } s, t \in \mathbb{R}\right\}$

*2. Criticize the following argument: By Exercise 1.1.13, for any vector $\mathbf{v}$, we have $0\mathbf{v} = \mathbf{0}$. So the first criterion for subspaces is, in fact, a consequence of the second criterion and could therefore be omitted.

♯3. Suppose $\mathbf{x}, \mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ and $\mathbf{x}$ is orthogonal to each of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Prove that $\mathbf{x}$ is orthogonal to any linear combination $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$.

4. Prove Proposition 3.2.

5. Given vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$, prove that $V = \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ is the *smallest* subspace containing them all. That is, prove that if $W \subset \mathbb{R}^n$ is a subspace and $\mathbf{v}_1, \dots, \mathbf{v}_k \in W$, then $V \subset W$.

♯6. (a) Let U and V be subspaces of $\mathbb{R}^n$. Define

$$U \cap V = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in U \text{ and } \mathbf{x} \in V\}.$$

Prove that $U \cap V$ is a subspace of $\mathbb{R}^n$. Give two examples.

(b) Is $U \cup V = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in U \text{ or } \mathbf{x} \in V\}$ a subspace of $\mathbb{R}^n$? Give a proof or counterexample.

(c) Let U and V be subspaces of $\mathbb{R}^n$. Define

$$U + V = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{u} + \mathbf{v} \text{ for some } \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}.$$

Prove that $U + V$ is a subspace of $\mathbb{R}^n$. Give two examples.

7. Let $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ and let $\mathbf{v} \in \mathbb{R}^n$. Prove that

$$\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}) \iff \mathbf{v} \in \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k).$$

♯*8. Let $V \subset \mathbb{R}^n$ be a subspace. Prove that $V \cap V^\perp = \{\mathbf{0}\}$.

♯9. Suppose $U, V \subset \mathbb{R}^n$ are subspaces and $U \subset V$. Prove that $V^\perp \subset U^\perp$.

♯10. Let $V \subset \mathbb{R}^n$ be a subspace. Prove that $V \subset (V^\perp)^\perp$. Do you think more is true?

♯11. Suppose $V = \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \subset \mathbb{R}^n$. Show that there are vectors $\mathbf{w}_1, \dots, \mathbf{w}_k \in V$ that are mutually orthogonal (i.e., $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ whenever $i \neq j$) that also span V. (Hint: Let $\mathbf{w}_1 = \mathbf{v}_1$. Using techniques of Section 2, define $\mathbf{w}_2$ so that $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$ and $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$. Continue.)

12. Suppose U and V are subspaces of $\mathbb{R}^n$. Prove that $(U + V)^\perp = U^\perp \cap V^\perp$. (See the footnote on p. 21.)

4 LINEAR TRANSFORMATIONS AND MATRIX ALGEBRA

We are heading toward calculus and the study of functions. As we learned in the case of one variable, differential calculus is based on the idea of the best (affine) linear approximation of a function. Thus, our first brush with functions is with those that are linear.

[2] In fact, both this definition and Proposition 3.2 work just fine for any sub*set* $V \subset \mathbb{R}^n$.